

# Review of: "first-workflow-Playarists v1"

Olga Pagnotta<sup>1</sup>

<sup>1</sup> University of Bologna

**Potential competing interests:** No potential competing interests to declare.

first-workflow-Playarists describes a clear overview on the process and steps that the researchers intend to follow for the analysis described in the abstract.

As for some suggestions, in 1.1 you propose two different approaches for analysing the ERIH-PLUS dataset relating to OpenCitations Meta. I would suggest to opt for the second approach, given the limited customisation in terms of use offered by SPARQL endpoint of OCMeta. The same for what concerns the fourth point in your workflow.

In addition, you refer to Meta as "OC dump". That could be confusing, because OpenCitations (note the plural, you refer to it sometimes as OpenCitation) offers various datasets and dumps to download. I would suggest you to refer to it either as Meta or, better, OCMeta.

In 1.2, you state that if a DOI cannot be found in DOAJ you assume that its Open Access status is False. I would suggest instead to add a column "unknown" in order to be as clear as possible about the real status of that publication.

In 2.1 I would suggest to give more information about the functioning of the script you are going to write for creating the mapping table for the disciplines.

In 2.3 you only assess the case in which missing information can be retrieved by looking at DOAJ. What happens if these metadata cannot be found there?

As a last point, I would suggest to provide a comprehensive visualisation to allow a clearer overall view of the situation.